Pak. J. Statist.
2017 Vol. 33(2), 75-93

ON A MULTIVARIATE EXPONENTIAL DISTRIBUTION BASED ON A D-VINE COPULA

Samia A. Adham[1] and **Azza I. Alzahrani**[2]
Department of Statistics, Faculty of Science, King Abdul-Aziz University,
P.O. Box 400, Jeddah 21411, Saudi Arabia.
Email: [1]sayadham@hotmail.com
[2]aialzahrany@kau.edu.sa

ABSTRACT

Exponential distributions have a pivotal role in the domain of reliability and life testing. The purpose of this paper is to construct a new multivariate exponential distribution based on the D-vine copula. Frank's copula from the Archimedean family and Gaussian copula from the elliptical family are implemented on the exponential distribution in the first stage. Then, the D-vine copula is applied in the second stage to construct a new multivariate exponential distribution. This paper focuses on Frank's and Gaussian copula and aimed at assessing the behavior of the association of the new multivariate distribution-parameters.

KEYWORDS

Multivariate Distribution, Bivariate Lifetime Copula, D-Vine, Simulation Study, Exponential Distribution

1. INTRODUCTION

Real-world phenomena are complex processes and their analysis requires certain assumptions to be made by the analyst. The purpose is to simplify the process of mathematical and statistical analysis. However, a balance needs to be maintained between simplifying assumptions and the reliability of the conclusions drawn based on those assumptions. There cannot be too many assumptions such that the reliability of the probabilistic model becomes questionable. The assumptions should be enough to facilitate calculations but not so many that the calculated model fails to represent the real world. In the decade of 50s, scientists initiated data analysis of the instrument's operating time. The aim of the study was to find a model that could resemble closely with the real world and has minimal assumptions. The findings of the study showed that the instrument's lifetime has exponential distribution (Gupta, Zeng and Wu, 2010). Further studies showed that exponential distribution is a useful tool for obtaining a first approximation. Hence, exponential distribution, its characterizations, properties, and models became an important area of study in the academic research. Some of the examples of the implementation of exponential distributions include time elapsed between spinal cords' impulses and calculation of telephone calls' length.

Construction of multivariate distributions is one of the classical fields of research in statistics, and hence, it continues to be an active field of research. Multivariate exponential distributions have proven to be an important class of distributions, and flexible multivariate distributions are necessary in many fields. A copula is a method of formalizing the dependence structures of random vectors. In many application fields, copula models have become increasingly popular during the last 10 years. Furthermore, in many cases of statistical modeling, it is essential to obtain the joint distribution between multiple random variables.

Although the marginal distribution of each of the random variables is known, their joint distribution may not be easy to obtain from these marginal distributions. Knowing the scale-free measures of dependence between random variables, the author used copula models to obtain the joint distribution. An interesting property of the copula method is that it is an approach to constructing a multivariate distribution of known marginal distributions.

However, the use of copula models in higher dimensions is considered a difficult challenge for researchers because constraints exist on the parameters of multivariate copulas, which can create inflexibility. The first appearance of pair-copulas was also noted(Joe, 1996). Two research studies have introduced the main concept of pair-copula constructions into a cascade of bivariate copulas and presented a method of organizing this process through vines using a graphical model; these researchers also briefly discussed simulations from vines (Bedford & Cooke, 2001; Kurowicka & Cooke, 2005). To overcome associated problems, a study developed the pair-copula constructions (PCCs) and reviewed easy and clear methods from algorithms to simulate the execution of D-vines and C-vines (Aas, et al., 2009). The D-vine copula and its inference were presented by a study in their investigational process (Aas& Berg, 2009). A study discussed the estimation of the parameters of PCCs, including the stepwise semi-parametric (SSP) estimator. Also, he presented its asymptotic properties. This study further selected the most suitable pair copula for the PCCs in a given data set. Moreover, he presented a multivariate exponential distribution based on D-vine with pair-copula: Gumbel copula for all cases. He also introduced estimated parameters for D-vine only (Haff, 2013).

The canonical vine (C-vine) and D-vine are popular types of PCCs, and the D-vine is more easily applied and more flexible than the C-vine. The C-vine relationship for one variable was observed to be predefined. Nevertheless, the selection of pairs for modeling the dependence in D-vine is unconstrained process.

The process of globalization has created a close interconnection among the economies of the world. Due to this close connectivity, fluctuations in one financial market have a substantial impact on other markets of the world. It was recently witnessed when China devalued its currency. Almost all stock markets of the world were affected from these changing economic conditions in China. Similarly, reduction in oil prices affected the markets across the world. All these factors necessitate understanding the relevance among different markets to minimize the risks of investment. The traditional method applied in this respect is the Pearson correlation. However, this type of relationship is relevant only when linear relationships are involved between two variables. In the real world, the distributions of financial variables are neither symmetric nor linear. The researchers stress the need of a new method that could evaluate the correlations. Copula functions have this capability that they can analyze the variables that are asymmetric and nonlinear in nature (Chen, Yang & Zhou, 2012). Another important feature of copula functions is that marginal distributions in these functions are unrestricted. Vine copulas are preferred for their flexibility of dealing with complex dependence patterns. These dependencies can be evaluated in a tree structure that provides ease in the evaluation of multiple dependencies (Allen, et al. 2013).

In this study, the author focuses on the D-vine copula. The paper focuses on Frank's and Gaussian copulas because both are symmetric; however the classes are different. The author is interested to see whether the copulas family is affecting on the exponential distribution or not.

The remainder of the paper is organized as follows. Section 2 describes the D-vine copula and h-function and explains the theoretical inference of the D-vine copula. Section 3 introduces construction of the new multivariate exponential distribution based on the D-vine copula. Section 4 describes a simulation study performed on the proposed multivariate exponential distribution based on the D-vine copula. Finally, the results are discussed in the final section together with conclusions.

2. THE D-VINE COPULA

Copula modeling techniques have gained increased importance during the last decade. The estimation and modeling methods were first presented by Sklar in 1959 (Flores, 2009). The unique feature of this technique is that it permits the decomposition of the distribution. The multivariate distribution is decomposed and univariate margins are formed (Kauermann & Schellhase, 2014). These univariate margins are exhibited through copula. According to the theorem of Sklar, the joint distribution can be expressed as (Flores, 2009):

$$F(x_{1 \ldots\ldots\ldots} x_p) = C\left(F_1(x_1) \ldots . F_p(x_p)\right)$$

In the above equation, $x_1, x_2, x_3, \ldots., x_p$ represent p-dimensional random vector and $F_1(x_1), F_2(x_2), F_3(x_3), \ldots, F_p(x_p)$ are univariate marginal distributions. The distribution function is denoted by C, the copula.

A study emphasize the usefulness of copulas in case of multivariate vector for the modeling of dependence (Haug, Kluppelberg & Peng, 2011). They stressed the need of testing and estimation methods for tail copulas and extreme value copulas. In the literature, there are three types of vine copulas called C-vine, D-vine and R-vine copulas. These regular vine copulas present a specific method used to decompose multivariate densities. High dimensional distributions can be constructed by applying the regular vines of the PCCs. In this paper, the author focuses on the D-vine copula (for additional details, Section 4). Figure 1, shows the D-vine specifications in three dimensions. For three variables, there are two treesT_j. Tree T_j has $d+1-j$ nodes and $i=d-j$ edges, where $j=1, 2$ and d=number of dimensions. In a pair-copula such as that used to build the modeling dependence, the author finds that each edge is associated with two variables. Furthermore, the dependencies between variables represent the naming of the edge in the connecting pair-copula density. In the first tree T_1, the dependence between two pairs of variables, (1, 2), (2,3) are used to build the corresponding pair-copula, $c_{1,2}(.;\theta_{1,2})$, $c_{2,3}(.;\theta_{2,3})$. In the second tree, T_2, the third variable (1,3|2) is modeled by using the conditional dependence with the associated pair-copula density $c_{1,3|2}(.;\theta_{1,3|2})$. Where $c_{1,2}, c_{2,3}, c_{1,3|2}$ as the densities of copula $C_{1,2}, C_{2,3}, C_{1,3|2}$.

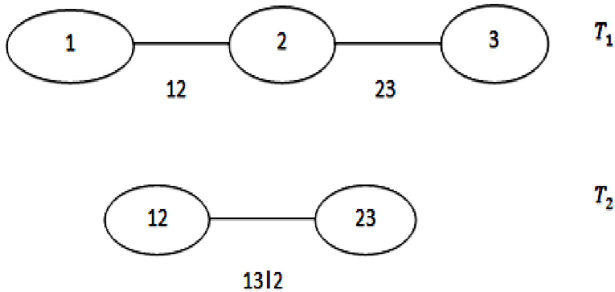

Fig. 1: Tree Representation of Three-Dimensional D-vine

According to the above explanation of the D-vine specification, one can write the multivariate density function of *d*-dimension based on the D-vine copula as follows

$$f(x_1,\dots,x_n)=\prod_{k=1}^{n} f(x_k)\prod_{j=1}^{n-1}\prod_{i=1}^{n-j} c_{i,i+j|i+1,\dots,i+j-1}\left(F\left(x_i|x_{i+1},\dots,x_{i+j-1}\right),F\left(x_{i+j}|x_{i+1},\dots,x_{i+j-1}\right)\right), \tag{1}$$

The representation in Eq. (1) has been suggested in a past study (Aas et al., 2009). In the three-dimensional case (*d*=3), this definition becomes

$$f(x_1,x_2,x_3)=f(x_1).f(x_2).f(x_3) .c_{1,2}\left(F(x_1),F(x_2)\right).c_{2,3}\left(F(x_2),F(x_3)\right) .c_{1,3|2}\left(F(x_1|x_2),F(x_3|x_2)\right). \tag{2}$$

According to both C-vine and D-vine copulas, one can choose among the three variables x_1, x_2 and $x_3$in six different ways. Hence, this paper is concerned with the D-vine copula, so there are three different ways, which provides the vine permutation property between random variables. It is now clear that each conditional marginal can be decomposed into the appropriate pair-copula times a conditional marginal density, using the general formula:

$$f(x|\mathbb{v})=c_{x,v_j|\mathbb{v}_{-j}}\left(F\left(x|\mathbb{v}_{-j}\right),F\left(v_j|\mathbb{v}_{-j}\right)\right).f\left(x|\mathbb{v}_{-j}\right), \tag{3}$$

where $\mathbb{v}=(\boldsymbol{v_1},\boldsymbol{v_2},\dots,\boldsymbol{v_n})'$and $\mathbb{v}_{-j}$denotes the vector $\mathbb{v}$, excluding the component v_j.

2.1 The *h*-function

The crucial question for constructing a PCCs model is how to obtain conditional distributions $F_{v_j|\mathbb{v}_{-j}}$ for a d-dimensional vector $\mathbb{v}$. The author uses the h-function to represent the conditional distribution function of the cumulative distributions. The notation of the h-function is introduced for convenience.

In Eq. (3), there is a need of an expression for the arguments of the $c_{x,v_j|\mathbb{v}_{-j}}$ arguments, i.e., $F_{x|\mathbb{v}_{-j}}$ and $F_{v_j|\mathbb{v}_{-j}}$. In (Joe, 1996) the following relationship is derived (under certain regularity conditions)

$$F_{x|\mathbb{v}} = \frac{\partial C_{x,v_j|\mathbb{v}_{-j}}}{\partial F_{v_j|\mathbb{v}_{-j}}}. \tag{4}$$

Components of the derivation of (4) are shown in the Appendix. If $\mathbb{v}$ is univariate, i.e., $\mathbb{v} = v$, Eq.(4) reduces to

$$F_{x|v} = \frac{\partial C_{xv}}{\partial F_v}. \tag{5}$$

However, when x and v are uniform, the expression in (5) is called the h-function. (See ref. 5).

$$h(x, v.\Theta) = F_{x|v} = \frac{\partial C_{x,v}(x,v,\Theta)}{\partial v}. \tag{6}$$

In Eq. (6), Θ is the vector of parameters for the current copula. The $h(.)$ is the conditioning variable. Hence, $h^{-1}(x, v, \Theta) = F_{x|v}^{-1}$ as the inverse of $h(x, v, \Theta)$ with respect to x. These examples are the most common and applicable copulas, and their h-functions.

2.2 Full Inference for a D-vine Copula

A study discussed the estimation parameters for PCCs(Haff, 2013). The D-vine log-likelihood with data set $\boldsymbol{x} = (x_{1,k}, \ldots, x_{n,k}), k = 1, \ldots, N; j = 1, \ldots, d)$ and parameter set θ is given by

$$l(\theta|\boldsymbol{x}) = \sum_{k=1}^{N}\sum_{i=1}^{d-1}\sum_{j=1}^{d-i} \log[c_{j,j+i|(j+1)\ldots(j+i-1)}(F_{j|(j+1),\ldots,(j+i-1)}, \\ F_{j+i|j+1,\ldots,j+i-1}|\theta_{j,j+i|(j+1),\ldots,j+i-1})]. \tag{7}$$

where the conditional distributions $F_{j|(j+1),\ldots,(j+i-1)}$ and $F_{j+i|j+1,\ldots,j+i-1}$ are determined using the h-function. Additionally, applying the previous relationship can raise a pair-copula term in tree j, which is possible using the pair-copulas for previous trees $1, \ldots, j-1$ and sequentially applying the relationship in Eq. (7). The log-likelihood should be numerically maximized over all parameters.

The root nodes should be determined in each tree. The dependence with respect to one variable occurs only in the first D-vine tree. The selection of copulas for bivariate variables is modeled in the same way as the first root node. The second root node is conditioned on the variable pairs. Generally, the root node selected for each tree and all dependence pairs take into account the modeled conditions for this node, i.e., for all root nodes in the trees, which is also in accordance with a past study (Nævestad, 2009).

In the domain of actuarial analysis, an important factor is the analysis of risk management. In this respect, it is important to take into account aggregate claims in Archimedean copula. Portfolio of single claims to generate the aggregate claims can be represented in the following equation (Furmanczyk, 2015):

$$S_n = \sum_{i=1}^{n} X_i$$

In the above equation, X_1, X_2, X_3, ...,X_n denote the single claims. The importance and uniqueness of this equation is that n dependent risks are represented by a portfolio. The overall dependence structure can then be represented by the Archimedean copula.

This work, applies the following sequential estimation procedure to estimate the values of the parameters that maximize the D-vine log-likelihood in Eq. (7):

Choose the first root node and determine all pair-copula types with respect to this root node in the first tree.

Select the copulas and the estimated parameters using the original data. Use the copula parameters that were rated in the first tree and the h-function condition as shown in the Appendix. Next, compute the observations (i.e., distribution functions conditioned on the first root node) for the second tree.

Choose the second root node and determine all pair-copula types with respect to this root node in the second tree. Use the conditional distribution function from (c) to estimate the parameters in the selected copulas in the second tree.

3. CONSTRUCTION OF THE NEW MULTIVARIATE EXPONENTIAL DISTRIBUTION

Assuming that a univariate random variable X follows an exponential distribution with parameter α, it is well known that the probability density and the cumulative distribution functions of X are respectively given by

$$f_X(x) = \alpha e^{-\alpha x}, x > 0;\ (\alpha > 0), \tag{8}$$

$$F_X(x) = 1 - e^{-\alpha x}, x > 0. \tag{9}$$

The concept of constructing the new multivariate exponential distribution based on PCCs can be demonstrated in two steps. The first step uses two distinct bivariate exponential distributions. In this step, the author applies Frank's and Gaussian copulas separately to the exponential distribution with the distribution function given by (8), considering different values of the parameter α. The second step applies the D-vine copula to the two bivariate exponential distributions obtained in the first step to construct the new multivariate exponential distribution. A study defines the parameters that may be used in case of exponential distributions for new characterizations (Bairamov, 2000). He emphasizes that the underlying distribution must either be new worse than used (NWU) or new better than used (NBU). Further details are given below.

Step 1:

Construct two bivariate exponential distributions based on Frank's and Gaussian copulas.

Let $\upsilon_j = F_{X_j}(x_j)$, where $F_{X_j}(x_j)$ is the distribution function of the exponential distribution, given by Eq.(9) after indexing X and α by j, $j = 1,2$. Next, the joint copula distribution functions of υ_1 and υ_2 based on Frank's copula take the following form (Nelsen 2010):

$$C(\upsilon_1, \upsilon_2; \theta) = -\theta^{-1} \log\left\{ 1 + \frac{\left(e^{-\theta\upsilon_1} - 1\right)\left(e^{-\theta\upsilon_2} - 1\right)}{e^{-\theta} - 1} \right\}, \tag{10}$$

where θ as a dependence parameter that may assume any real value $\{(-\infty, \infty)/0\}$. Hence, the joint distribution functions of the two exponential random variables $X_1$and X_2 are

$$F_U(u_1, u_2) = C(\upsilon_1, \upsilon_2, \theta),$$

where $C(\upsilon_1, \upsilon_2; \theta)$ is given by Eq.(10). Consequently, the bivariate exponential density function based on Frank's copula of the pair (X_1, X_2) is given by

$$f_{X_1,X_2}(x_1, x_2) = f_{X_1}(\mathrm{x}_1) f_{X_2}(x_2) c'(\upsilon_1, \upsilon_2), \tag{11}$$

where

$$\mathrm{C}'(\upsilon_1, \upsilon_2) = \frac{\partial^2 \mathrm{C}(\upsilon_1, \upsilon_2; \theta)}{\partial\upsilon_1 \partial\upsilon_2} = \frac{\theta \exp\{\theta(\upsilon_1 + \upsilon_2)\}\left(e^{\theta} - 1\right)}{\{\left(e^{\theta\upsilon_1} - 1\right)\left(e^{\theta\upsilon_2} - 1\right) + e^{\theta} - 1\}^2}, \theta \neq 0, \tag{12}$$

and $\mathrm{f}_{\mathrm{X_j}}\left(x_{\mathrm{j}}\right)$ is the probability density function of the exponential distribution given by (8) after indexing X and α by $j = 1,2$.

Similarly, the Gaussian copula is implemented in the exponential distribution margins to obtain a bivariate exponential distribution based on the Gaussian copula. The Gaussian copula with association parameter ρ as presented (Nelsen, 2010) is,

$$\mathrm{C}(\upsilon_1, \upsilon_2; \rho) = \Phi_\rho\left(\Phi^{-1}(\upsilon_1), \Phi^{-1}(\upsilon_2)\right), \tag{13}$$

where Φ and Φ^{-1} are the $N(0,1)$ distribution function and its inverse, i.e., $\Phi^{-1}\{\Phi(x)\} = x$, and

$$\Phi(x) = \int_{-\infty}^{x} \frac{\exp(\frac{-z^2}{2})}{\sqrt{2\pi}} dz,$$

and Φ_ρ is the bivariate standard normal distribution function with ρ correlation given by

$$\Phi_\rho(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} \frac{\exp\{\frac{-1}{2(1-\rho^2)}(z_1^2 - 2\rho z_1 z_2 + z_2^2)\}}{2\pi\sqrt{1 - \rho^2}} dz_2 dz_1.$$

Step 2:

Implement the D-vine to construct the new multivariate exponential distribution.

Let $\underline{X} = (X_1, X_2, X_3) \sim F_{\underline{X}}(\underline{x})$, where $F_{\underline{X}}(\underline{x}) = C(F_{X_1}(x_1), F_{X_2}(x_2), F_{X_3}(x_3))$, if the joint distribution function has marginals $F_{X_j}(x_j)$ and corresponding densities $f_{X_j}(x_j)$ and $j = 1,2,3$. Hence, by recursive conditioning, the joint density function is defined using

$$f_{\underline{X}}(\underline{x}) = f_{X_1,X_2,X_3}(x_1, x_2, x_3) = f_{X_1}(x_1) f_{X_2|X_1}(x_2|x_1) f_{X_3|X_1,X_2}(x_3|x_1, x_2). \quad (14)$$

Consequently,

$$f_{X_2|X_1}(x_2|x_1) = \frac{f_{X_2,X_1}(x_1, x_2)}{f_{X_1}(x_1)} = \frac{c_{1,2}\left(F_{X_1}(x_1), F_{X_2}(x_2)\right) f_{X_1}(x_1) f_{X_2}(x_2)}{f_{X_1}(x_1)}$$
$$= c_{1,2}\left(F_{X_1}(x_1), F_{X_2}(x_2)\right) f_{X_2}(x_2),$$

and

$$f_{X_3|X_1,X_2}(x_3|x_1, x_2) = \frac{f_{X_1,X_3|X_2}(x_1, x_3|x_2)}{f_{X_1|X_2}(x_1|x_2)}$$
$$= \frac{c_{1,3|2}(F_{X_1|X_2}(x_1|x_2), F_{X_3|X_2}(x_3|x_2)) f_{X_1|X_2}(x_1|x_2) f_{X_3|X_2}(x_3|x_2)}{f_{X_1|X_2}(x_1|x_2)}$$
$$= c_{1,3|2}(F_{X_1|X_2}(x_1|x_2), F_{X_3|X_2}(x_3|x_2)) f_{X_3|X_2}(x_3|x_2)$$
$$= c_{1,3|2}(F_{X_1|X_2}(x_1|x_2), F_{X_3|X_2}(x_3|x_2)) c_{2,3}\left(F_{X_2}(x_2), F_{X_3}(x_3)\right) f_{X_3}(x_3).$$

From Eq. (14), one can find the probability density function that can be represented with the bivariate copula models $C_{1,2}, C_{2,3}$ and$C_{1,3|2}$ with densities $c_{1,2}, c_{2,3}$ and $c_{1,3|2}$, so-called pair-copulas, which may be chosen independently, and thus, we can achieve a wide range of different structures of dependence. Because of the decomposition in Eq. (14), there exist many such iterative PCCs, which are not unique, as stated in the notes(Kao, 2011). The probability density function for a new multivariate exponential distribution based on the D-vine in the three-dimensional case ($p = 3$) is given by:

$$f_{\underline{X}}(\underline{x}) = f_{X_1,X_2,X_3}(x_1, x_2, x_3) = f_{X_1}(x_1). f_{X_2}(x_2). f_{X_3}(x_3)$$
$$. c_{1,2}\left(F_{X_1}(x_1), F_{X_2}(x_2)\right). c_{2,3}\left(F_{X_2}(x_2), F_{X_3}(x_3)\right)$$
$$. c_{1,3|2}\left(F_{X_1|X_2}(x_1|x_2), F_{X_3|X_2}(x_3|x_2)\right) \quad (15)$$

where $c_{1,2}$ is the density of the bivariate exponential based on Frank's copula given by Eq.(10), $c_{2,3}$is the density of the bivariate exponential based on the Gaussian copula given by Eq.(13), and $c_{1,3|2} = f_{X_1,X_3|X_2}(x_1, x_3|x_2)$, the conditional distribution.

4. SIMULATION STUDY

By simulating samples of sizes 50, 100, and 1000, to simulate a data set from a specified three-dimensional D-vine copula, for simplicity, it was assumed that the marginal distributions are uniformly drawn from exponential distributions. Next, the author generates the three-dimensional sample $\underline{x} = (x_1, x_2, x_3)$ from a specified D-vine copula as follows:

1. Assume that w_1, w_2, w_3 are independent uniform [0,1] random variables and proceed with

$$x_1 = w_1,$$

$$x_2 = F_{2|1}^{-1}(w_2|x_1),$$

$$x_3 = F_{3|12}^{-1}(w_3|x_1, x_2).$$

The conditional distribution functions $F_{j|i}\left(x_j \middle| \underline{x}_{-j}\right)$, where $j = 1,2,3, \underline{x}_{-j}$ is the vector $\underline{x}$ after eliminating x_j, are given by

$$F_{2|1}(x_2|x_1) = \frac{\partial C_{12}(F_2(x_2), F_1(x_1)|\theta_{12})}{\partial F_1(x_1)},$$

$$F_{3|12}(x_3|x_1, x_2) = \frac{\partial C_{23|1}(F_{3|1}(x_3|x_1), F_{2|1}(x_2|x_1)|\theta_{23|1})}{\partial F_{2|1}(x_2|x_1)},$$

Using the relationship between the conditional distribution function and the h-function in Subsection (2.1), we obtain

$$F_{2|1}(x_2|x_1) = h_{12}\,(x_2|x_1, \theta_{12}),$$

and

$$\begin{aligned} F_{3|12}(x_3|x_1, x_2) &= h_{23|1}(F_{3|1}(x_3|x_1)|F_{2|1}(x_2|x_1), \theta_{23|1}) \\ &= h_{23|1}\left(h_{13}(x_3|x_1, \theta_{13})\middle|h_{12}(x_2|x_1, \theta_{12}), \theta_{23|1}\right). \end{aligned}$$

2. Set $x_1 = w_1$,

$$x_2 = h_{12}^{-1}(w_2|x_1, \theta_{12}),$$
$$x_3 = h_{13}^{-1}\left(h_{23|1}^{-1}\left(w_3\middle|h_{12}(x_2|x_1, \theta_{12}), \theta_{23|1}\right)\middle|x_1, \theta_{13}\right),$$

where the h-function and its inverse h^{-1} are given in Appendix A.

For the data simulation the authorized the package CD-vine in R packages (Brechmann & Schepsmeier, 2013). Furthermore, the Vine package implemented by a study was also applied (Fernandez & Soto, 2014). The marginal exponential distribution parameters are estimated using the maxLik package (details in Toomet et al., 2013).

Three random samples each is of size 100 are simulated from Frank's and Gaussian copulas, separately, with different values for the copula parameters(i.e. a different sample is simulated for each parameter). The estimated mean and mean square error (MSE) of the simulated samples are computed. Table 1 and Table 2 list different parameter values of Frank's copula and Gaussian copula, respectively. They include the means and the MSE of the generated samples. The simulated observations with the smallest MSE, is chosen to ensure the homogeneity. i.e., for Frank's and Gaussian copula parameters, the author selected the values 5 and 0.75 respectively. It must be noted that the zero is not included in the range of Frank's copula parameter.

Table 1
Estimated Mean and MSE of Simulated Samples with Different Values of Frank's Copula Parameter

Parameter	Mean est	MSE
7	6.6178	0.0678
5	4.6795	0.0508
-5	-5.1878	0.0534
-7	-7.1510	0.0726

Table 2
Estimated Mean and MSE of Simulated Samples with Different Values of the Gaussian Copula Parameter

Parameter	Mean est	MSE
0.75	0.7377	0.0001
0.5	0.4707	0.0004
0	-0.0444	0.0009
-0.5	-0.5183	0.0004
-0.75	-0.755	0.0001

For estimating the parameters of the new multivariate exponential distribution, a random vector $\underline{X} = (X_1, X_2, X_3)$ is simulated assuming that each X_i, *i=1,2,3*, is a standard exponential (i.e., X_i follows an exponential distribution with parameter $\alpha_i = 1$).

In Step 1, the cdfs of X_1 and X_2 are applied to construct one bivariate exponential based on Frank's copula. Similarly, the cdfs of X_2 and X_3 are applied to construct another bivariate exponential based on the Gaussian copula.

Next, random samples of sizes N=50, 100, 1000 are simulated from the bivariate exponential distribution based on Frank's copula with $\theta_1 = 5$. Similarly, samples are simulated from the bivariate exponential distribution based on the Gaussian copula with $\theta_2 = 0.75$.

The CDVine package in R is applied to estimate the copula parameters θ_1 and θ_2. Table 3 shows the maximum likelihood estimates (MLEs) of Frank's and Gaussian copula parameters θ_1 and θ_2.

Table 3
MLEs of Frank's and Gaussian Copula Parameters $\boldsymbol{\theta_1}$ and $\boldsymbol{\theta_2}$

Sample Size: N	Initial value	50		100	1000
Frank	5	$\widehat{\theta_1}$	5.162	5.526	4.937
Gaussian	0.75	$\widehat{\theta_2}$	0.777	0.7622	0.785

Next, the maxLik package in R is applied to compute the MLE of the marginal parameters of each of Frank's and Gaussian bivariate exponential distributions. Table 4 shows the MLEs of the marginal parameters of the two bivariate exponential distributions, where the parameter values are $\alpha_1 = \alpha_2 = \alpha_3 = 1$.

Table 4
Estimated Parameters for the Bivariate Exponential-Distribution-based Copula

Sample Size: N	Frank's		Gaussian	
	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_2$	$\hat{\alpha}_3$
50	1.14	1.04	0.98	0.97
100	0.95	0.95	1.08	1.08
1000	1.08	1.08	1.08	1.08

It can be observed that the ML estimates of the parameters θ_1, θ_2 and α_i, i=1,2,3, are close to their actual values. Moreover, from Tables 3 and 4, it is clear that the sample sizes do not affect the estimation (see, e.g., Flores, 2009).

Figures 2 and 3 present the cdf and pdf plots of the bivariate exponential distributions based on Frank's and Gaussian copulas, respectively.

Fig. 2: Plots of cdf and pdf the Bivariate Exponential Distribution based on Frank's Copula

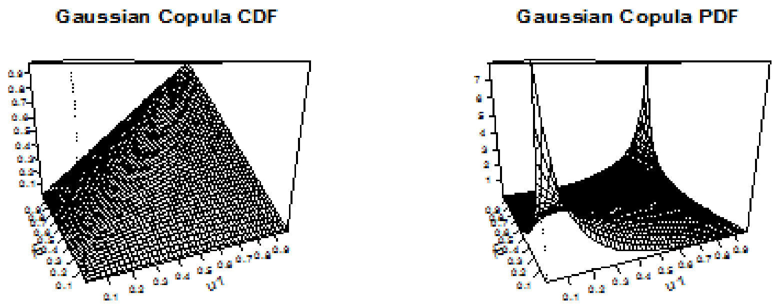

Fig. 3: Plots of cdf and pdf the Bivariate Exponential Distribution based on the Gaussian Copula

In step 2, A D-vine copula is simulated; the new multivariate exponential distribution. Random samples of sizes N=50, 100, 1000 are simulates from the cumulative exponential distributions for a three-dimensional D-vine copula. Frank's and Gaussian copulas are applied as described in Section 3. The parameters of the new multivariate exponential distribution based on the proposed three-part model using different parameter marginal distributions are estimated.

4.1 Selection among D-vine Copula Models

The vector entrances in the D-vine should be compatible with the pairs and the associations between the selections of copula pair terms (Brechmann & Schepsmeier, 2013).

Table 5
Selection among D-vine Copula Models

(1,2),(2,3)	(Tree 1)
(1,3|2)	(Tree 2)

By applying an example of the D-vine copula model in three closed dimensions to a selection of pair-copula terms $c_{1,2}, c_{2,3}$, and $c_{1,3|2}$, the author finds the following. In tree 1, the pair-copulas $c_{1,2}$ and $c_{2,3}$ are Frank's copula with parameter $\theta_{1,2} = 5$ and Gaussian copula with parameter $\theta_{2,3}$ = 0.7, respectively. While in tree 2, pair copula $c_{1,3|2}$ in the tree 2 is a Gaussian copula with parameter $\theta_{1,3|2}$. Table 6 shows the settings for simulating the D-vine data for three different models. Thedatasetsof sample sizes 50, 100 and 1000 are generatedfrom this D-vine using the simulation algorithm presented previously.

The sequential estimation procedure in Subsection 2.2 is applied to select a D-vine model and verify whether this D-vine model is consistent with the settings. Then the pair-copula is chosen using the AIC. The selected copula model for each pair is found to be consistent with the settings. Table 7 shows the computed values of the AICs for each pair-copula. The smallest AIC indicates a better estimate. Hence, the computed values of the AICs in Table 7 suggest that the three models of the D-vine distribution with Gaussian copulas and a parameter equals to 0.5 provide better fit for the data than other models.

Although the sequential estimation usually provides good parameter estimates, these estimates can be improved by a joint MLE. The numerical results for all estimated parameters for the pair-copula are presented in Table 8.

Table 6
Settings for Simulation of D-vine Data

	Pairs	**Family**	**Parameter**		
			N=50	**N=100**	**N=1000**
Model (1) $\alpha_1 = (1,1.5,2)$	$c_{1,2}$	Frank	$\theta_{1,2=}5$	$\theta_{1,2=}5$	$\theta_{1,2=}5$
	$c_{2,3}$	Gaussian	$\theta_{2,3} = 0.7$	$\theta_{2,3} = 0.7$	$\theta_{2,3} = 0.7$
	$c_{1,3|2}$	Gaussian	$\theta_{1,3|2} = 0.45$	$\theta_{1,3|2} = 0.58$	$\theta_{1,3|2} = 0.56$
Model(2) $\alpha_2 = (0.5,1,1.5)$	$c_{1,2}$	Frank	$\theta_{1,2} = 5$	$\theta_{1,2} = 5$	$\theta_{1,2} = 5$
	$c_{2,3}$	Gaussian	$\theta_{2,3=}0.7$	$\theta_{2,3=}0.7$	$\theta_{2,3=}0.7$
	$c_{1,3|2}$	Gaussian	$\theta_{1,3|2} = 0.50$	$\theta_{1,3|2} = 0.53$	$\theta_{1,3|2} = 0.54$
Model (3) $\alpha_3 = (1,1,1)$	$c_{1,2}$	Frank	$\theta_{1,2} = 5$	$\theta_{1,2} = 5$	$\theta_{1,2} = 5$
	$c_{2,3}$	Gaussian	$\theta_{2,3} = 0.7$	$\theta_{2,3} = 0.7$	$\theta_{2,3} = 0.7$
	$c_{1,3|2}$	Gaussian	$\theta_{1,3|2} = 0.52$	$\theta_{1,3|2} = 0.51$	$\theta_{1,3|2} = 0.58$

Table 7
AICs for each Pair-Copula

Pairs		Frank	Gaussian	Gaussian
Model (1)	N=50	-357.48	-455.75	-235.06
Model (2)		-378.59	-519.30	-162.94
Model (3)		-747.53	-822.79	-410.49
Model (1)	N=100	-31.67	-46.02	-14.42
Model (2)		-16.28	-39.48	-22.60
Model (3)		-64.73	-76.86	-26.59
Model (1)	N=1000	-12.79	-36.25	-5.15
Model (2)		-7.25	-27.92	-11.78
Model (3)		-30.53	-32.30	-13.63

Table 8
Estimated Parameters for a Three-Dimensional D-vine Copula

Parameter		N=50		N=100		N=1000		Real Value
		Start	Final	Start	Final	Start	Final	
Model 1	$\boldsymbol{\theta_{1,2}}$	4.2951	4.5434	5.5784	5.6837	5.5258	5.6484	5
	$\boldsymbol{\theta_{2,3}}$	0.6647	0.8235	0.7122	0.7914	0.7449	0.7467	0.7
	$\boldsymbol{\theta_{1,3\|2}}$	0.4927	0.4054	0.4832	04387	0.5170	0.5130	0.5
	Log-likelihood	33.433		51.980		534.069		
Model 2	$\boldsymbol{\theta_{1,2}}$	3.9908	5.1626	3.442	4.5545	4.724	4.814	5
	$\boldsymbol{\theta_{2,3}}$	0.6844	0.7748	0.304	0.6639	0.4056	0.7003	0.7
	$\boldsymbol{\theta_{1,3\|2}}$	0.4419	0.6206	0.5262	0.4901	0.459	0.4235	0.5
	Log-likelihood	29.4409		45.0130		534.085		
Model 3	$\boldsymbol{\theta_{1,2}}$	4.5813	4.7845	5.1363	5.344	5.021	5.980	5
	$\boldsymbol{\theta_{2,3}}$	0.6498	0.6407	0.7518	0.7522	0.7004	0.7005	0.7
	$\boldsymbol{\theta_{1,3\|2}}$	0.5899	0.5908	0.5055	0.5023	0.5818	0.5826	0.5
	Log-likelihood	30.7301		78.039		807.238		

Figure 4 shows a bivariate contour plot of the Frank copula corresponding to a bivariate exponential distribution with different margins. In addition, the figure shows the specified bivariate Frank copula and parameter values.

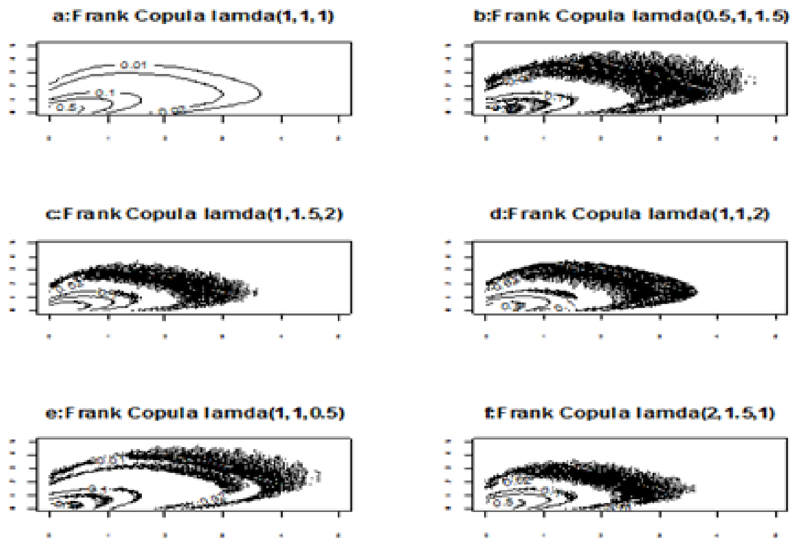

Fig. 4: Contour Plot of Frank's Copula for a Marginal Exponential Distribution

A study has proposed the vector of contour levels for exponential margins with default value levels = (0.01, 0.05, 0.1, 0.15, 0.2) as typically good choices (Brechmann & Schepsmeier, 2013). From Figure 5, it is noted that the probability density function takes different shapes. Thus, one can use the bivariate exponential distribution for the analysis of bivariate skewed data sets. This paper constructed the new multivariate exponential distribution based D-vine copula.

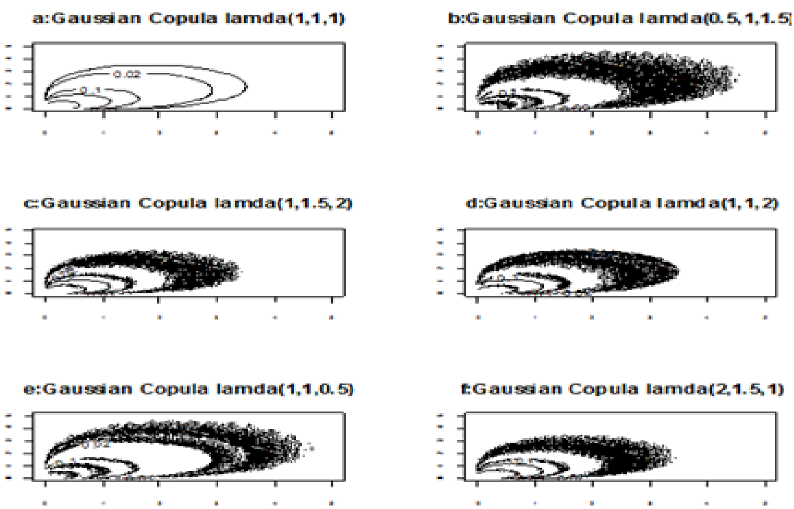

Fig. 5: Contour Plot of the Gaussian Copula for a Marginal Exponential Distribution

4.2 Simulation Results for the Estimated Parameters of the Marginal Distribution

The author considered three different sets of the marginal parameters: $\alpha_1 = (1,1.5,2)$, $\alpha_2 = (0.5,1,1.5)$, and $\alpha_3 = (1,1,1)$. The results are the estimated marginal parameters listed in Table 9. Finally, in the table, it is shown that the estimated parameter from the inverse function is close to the proposed parameter, meaning that the estimate is a good value.

Table 9
Estimated Parameter Marginal Distribution from the Maximum Likelihood Method

Parameter		$\widehat{\alpha_1}$	$\widehat{\alpha_2}$	$\widehat{\alpha_3}$
Models	True values	EV by MLE N=1000		
Model 1	(1,1.5,2)	1.013	1.601	2.075
Model 2	(0.5,1,1.5)	0.636	1.201	1.511
Model 3	(1,1,1)	0.974	1.045	1.039

Figure 6 shows the pair-wise scatter plots of the three variables as verification. The same plot is used for all of the pairs, meaning that there are no lower or upper limits. The distributions might be symmetric in the components, and the observations in the lower and upper plots could have equal values or none. The same figure shows the contour plot of the density functions of the copula for the pairs (1,2; 2,3; 1,3|2) and the estimated parameters from the simulated D-vine. The plots are all quite similar.

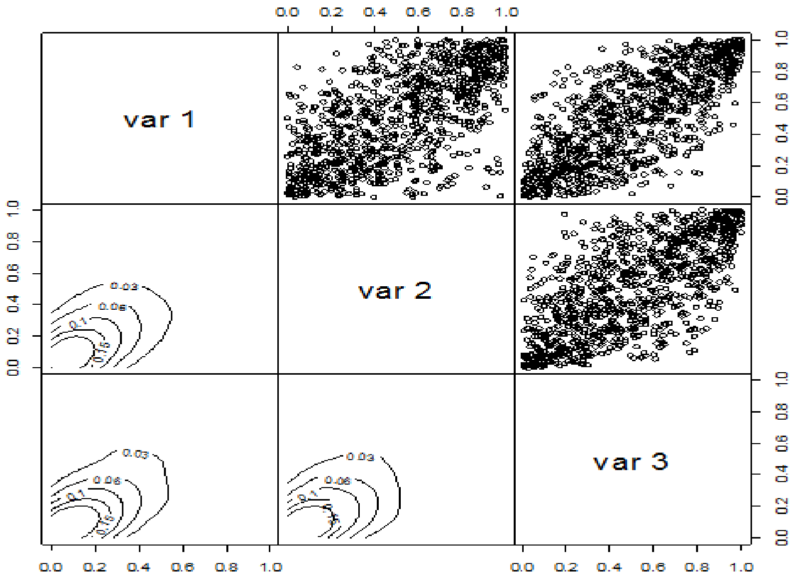

Fig. 6: Pair-wise Scatter Plots and Contours of the Three Variables

5. CONCLUSION

This article introduced the new multivariate exponential distribution based on D-vine copulas for dependence structures in multivariate exponential distributions and illustrated its performance on simulated data from a three-dimensional D-vine. All parameters were estimated. The estimates were notably close to the real parameter values in the cases of large sample sizes. Moreover, the new multivariate exponential distribution based on PCCs with three dimensions is symmetrical. However, if higher dimensions are applied, an asymmetric multivariate exponential distribution based on PCCs is possible. The paper focused on Frank's and Gaussian copulas because both are symmetric, but their classes are different. The author concludes that the PCCs provide more flexibility in constructing multivariate models than known copulas.

6. FUTURE WORK

Copula functions have proved very significant in the statistical analysis when the variables under consideration are asymmetric and nonlinear. There is a high volatility and increased uncertainty in various aspects of lives. Hence, the assumptions of normal distribution may have a substantial impact on the conclusions drawn from statistical analysis. The new multivariate exponential distribution, based on D-vine copulas, is presented in this study. It can become a useful resource for the analysts in the engineering, financial, and academic world. The findings of the study authenticated that the estimations were close to the real parameters. Hence, the model has the potential of handling nonlinearity and asymmetry of data.

In future work, the author intends to apply the method with D-vine or C-vine in three dimensions on a real data set. In addition, the author plans to use a new distribution with a more flexible dependence structures and compare it with the new multivariate exponential parameter estimated in this study.

7. ACKNOWLEDGEMENTS

The author(s) are very thankful to all the associated personnel in any reference that contributed in/for the purpose of this research.

8. REFERENCES

2. Aas, K. and Berg, D. (2009). Models for construction of multivariate dependence–a comparison study. *Eur. J. Finance*, 15, 639-659. doi:10.1080/13518470802588767
3. Aas, K., Czado, C., Frigessi, A. and Bakken, H. (2009). Pair-copula constructions of multiple dependence. *Ins. Math. Econ.*, 44, 182-198. doi:10.1016/j.insmatheco.2007.02.001.
4. Allen, D.E., Ashraf, M.A., McAleer, M., Powell, R.J. and Singh, A.K. (2013). Financial dependence analysis: applications of vine copulas. *Statistica Neerlandica*, 67, 403-435.
5. Bairamov, I.G. (2000). On the characteristic properties of exponential distribution. *Annals of the Institute of Statistical Mathematics*, 52, 448-458. doi:10.1023/A:1004109117386

6. Bedford, T. and Cooke, R.M. (2001). Probability density decomposition for conditionally dependent random variables modeled by vines. *Ann. Math. Artif. Intel.*, 32, 245-268. doi:10.1023/A:1016725902970
7. Brechmann, E.C. and Schepsmeier, U. (2013). Modeling dependence with C-and D-vine copulas: The R-package CDVine. *J. Stat. Softw*., 52, 1-27.
8. Chen, J., Yang, T. and Zhou, H. (2012). *Correlation Analysis among International Soybean Futures Markets Based on D-vine Pair-Copula Method. Advances in Computer Science and Information Engineering*, Springer Berlin Heidelberg.
9. Embrechts, P., Frey, R. and McNeil, A. (2005). *Quantitative risk management*. Princeton Series in Finance, Princeton.
10. Fernandez, Y. and Soto, M. (2014). *Package 'vines'*. http://github.com/ygf/vines.
11. Flores, A. (2009).Testing Copula Functions as a Method to Derive Bivariate Weibull Distributions. *American Political Science Association* (APSA), 3-6.
12. Furmanczyk, K. (2015). Archimedean copulas with applications to VaR estimation. *Stat Methods Appl*., doi:10.1007/s10260-015-0326-7
13. Gupta, A.K., Zeng, W. and Wu, Y. (2010). *Exponential Distribution*. Birkhauser Boston.
14. Haff, I.H. (2013).Parameter estimation for pair-copula constructions. *Bernoulli*, 19, 462-491.
15. Haug, S., Klüppelberg, C. and Peng, L. (2011). Statistical models and methods for dependence in insurance data. *Journal of the Korean Statistical Society*, 40(2), 125-139.
16. Joe, H. (1996). *Multivariate Models and Multivariate Dependence Concepts*, vol. 73. CRC Press, Boca Raton.
17. Kao, Y.Z. (2011). *Portfolio selection based on C-Vine pair-copula constructions*. Master's Thesis, National Central University, Nanjing.
18. Kauermann, G. and Schellhase, C. (2014). Flexible pair-copula estimation in D-vines using bivariate penalized splines. *Statistics and Computing*, 24, 1081-1100. doi:10.1007/s11222-013-9421-5.
19. Kurowicka, D. and Cooke, R.M. (2005). Distribution-free continuous bayesian belief nets. *Mod. Stat. Math. Met. Reliabil*., 10, 309-323.
20. Nævestad, M. (2009). *Multivariate Distributions through Pair-Copula Construction: Theory and Applications*. Masters thesis, Institutt for Matematiske Fag, Trondheim.
21. Nelsen, R.B. (2010). *An Introduction to Copula*. (2nd ed.). Berlin: Springer-Verlag.
22. Sklar, M. (1959). *Functions de repartition à n dimensions ethers marges*. University Paris.
23. Toomet, O. (2013). *Package' maxLik'*. http://www.maxLik.org.

APPENDIX

A. Expressions of the h- and h^{-1}f Functions of Various Bivariate Copulas

Gaussian Copula

The distribution function of the Gaussian copula in two dimensions with correlation parameter ρis given by

$$C_N(u,v;\rho) = \Phi_\rho\left(\Phi^{-1}(u),\Phi^{-1}(v)\right),$$

where Φ_ρ is the bivariate Gaussian function with correlation parameter ρ and Φ denotes the standard univariate normal distribution function with Φ^{-1} as the inverse. For this copula, the h and h^{-1}functions are

$$h_N(x,v;\rho) = \Phi\left(\frac{\Phi^{-1}(x) - \rho\Phi^{-1}(v)}{\sqrt{1-\rho^2}}\right),$$

$$h_N^{-1}(u,v;\rho) = \Phi\left(\Phi^{-1}(u)\sqrt{1-\rho^2} + \rho\Phi^{-1}(v)\right).$$

Frank's copula is described by

$$C_F(u,v,\theta) = -\frac{1}{\theta}\log\left(1 + \frac{(\exp(-\theta u)-1)(\exp(-\theta v)-1)}{\exp(-\theta)-1}\right), \theta \in \mathbb{R}.$$

$$h_F(u,v,\theta) = \frac{g_u g_v + g_u}{g_u g_v + g_1},$$

$$h_F^{-1}(u,v,\theta) = -\frac{1}{\theta}\ln\left\{1 + \frac{ug_1}{1+g_v(1-u)}\right\}, \text{where} g_y = e^{-\theta y} - 1.$$

The derivations of these formulas are given in (Aas et al. 2009) and (Nævestad, 2009).

B. Bivariate Copulas

The most stable theorem is the Sklar's theorem (Sklar, 1959) which shows the described dependence between variables in statistics. Furthermore, this concept formalizes the important role of copulas and establishes the connection between multivariate and univariate margins for distribution functions.

Sklar's Theorem.

Let margins $F_1, \dots, F_d$. From F, we obtain a d-dimensional distribution function. Thus, a copula C can exist such that for all $\underline{x} = (x_1, \dots, x_d)' \in \mathbb{R}^d$

$$F(x) = C(F(x_1), \dots, F(x_d)) \quad \text{(B.1)}$$

if $F_1, \dots, F_n$ are continuous and C is unique. Conversely, if $F_1, \dots, F_d$ are cumulative univariate distribution functions, then the function F(x) defined by Eq. (B.1) is a joint distribution function with margins $F_1, \dots, F_d$, and C is a copula.

In this section, we introduce certain famous bivariate copulas that were used in this paper as candidates of the pair-copula.

The functions $C(u_1, .., u_n)$, where $n = 1, \dots d$, are multivariate distribution functions with d-dimensional copulas that can be used to characterize the dependency between d random variables while allowing for arbitrary marginal distributions. We develop multivariate copulas using only bivariate copulas as building blocks, and therefore, we concentrate in this section on d = 2. In particular, the famous theorem of (Sklar, 1959) describes the connection between the marginals and the copula of the joint distribution (Joe, 1997). For this purpose, let F(·,·) denote a bivariate cdf with marginal cdfs F_1 and F_2, respectively; thus, there exists a two-dimensional copula cdf defined as

$$F(x_1, x_2) = C\big(F_1(x_1), F_2(x_2)\big)(B.2)$$

such that for all $(x_1, x_2), \in R^2$holds. For continuous F_1 and F_2, C(·, ·) is unique and is defined by

$$C(x_1, x_2) = F(F_1^{-1}(x_1), F_2^{-1}(x_2)).$$

The given density is $c(u_1, u_2) = \frac{\partial^2 C(u_1, u_2)}{\partial u_1 \, \partial u_2}$.

The most important and commonly used copulas in finance are the Gaussian types. Both belong to the class of elliptical copulas (for a precise definition, see, e.g., (Joe, 1996). Another class that is often discussed and utilized is the Archimedean copulas (see, e.g., (Sklar, 1959)).